GENERALIZED CHEBYSHEV BOUNDS VIA SEMIDEFINITE PROGRAMMING*

LIEVEN VANDENBERGHE[†], STEPHEN BOYD[‡], AND KATHERINE COMANOR[§]

Abstract. A sharp lower bound on the probability of a set defined by quadratic inequalities, given the first two moments of the distribution, can be efficiently computed using convex optimization. This result generalizes Chebyshev's inequality for scalar random variables. Two semidefinite programming formulations are presented, with a constructive proof based on convex optimization duality and elementary linear algebra.

Key words. Semidefinite programming, convex optimization, duality theory, Chebyshev inequalities, moment problems.

AMS subject classifications. 90C22, 90C25, 60-08.

1. Introduction.

Chebyshev inequalities give upper or lower bounds on the probability of a set based on known moments. The simplest example is the inequality

$$\mathbf{Prob}(X < 1) \geq \frac{1}{1+\sigma^2},$$

which holds for any zero-mean random variable X on $\mathbf{R}$ with variance $\mathbf{E}\,X^2 = \sigma^2$. It is easily verified that this inequality is sharp: the random variable

$$X = \begin{cases} 1 & \text{with probability } \sigma^2/(1+\sigma^2) \\ -\sigma^2 & \text{with probability } 1/(1+\sigma^2) \end{cases}$$

satisfies $\mathbf{E}\,X = 0$, $\mathbf{E}\,X^2 = \sigma^2$ and $\mathbf{Prob}(X < 1) = 1/(1+\sigma^2)$.

In this paper we study the following extension: given a set $C \subseteq \mathbf{R}^n$ defined by strict quadratic inequalities,

$$C = \{x \in \mathbf{R}^n \mid x^T A_i x + 2b_i^T x + c_i < 0,\ i = 1, \ldots, m\}, \tag{1.1}$$

find the greatest lower bound on $\mathbf{Prob}(X \in C)$, where X is a random variable on $\mathbf{R}^n$ with known first and second moments $\mathbf{E}\,X$ and $\mathbf{E}\,XX^T$. We will see that the bound, and a distribution that attains it, are readily obtained by solving a convex optimization problem.

History. Several generalizations of Chebyshev's inequality were published in the 1950s and 1960s. We can mention in particular a series of papers by Isii [Isi59, Isi63, Isi64] and Marshall and Olkin [MO60], and the book by Karlin and Studden [KS66, Chapters XII-XIV]. Isii [Isi64] notes that Chebyshev-type inequalities can be derived using the duality theory of infinite-dimensional linear programming. He considers the problem of computing upper and lower bounds on $\mathbf{E}\,f_0(X)$, where X is a random variable on $\mathbf{R}^n$, whose distribution satisfies the moment constraints

$$\mathbf{E}\,f_i(X) = a_i, \quad i = 1, \ldots, m,$$

* Research supported in part by NSF under grants ECS-0524663 and ECS-0423905, ARO under ESP MURI Award No. DAAD19-01-1-0504, and AFOSR under grant AF F49620-01-1-0365.

[†]Electrical Engineerng Department, University of California, Los Angeles (vandenbe@ucla.edu).

[‡]Department of Electrical Engineering, Stanford University (boyd@stanford.edu).

[§]The RAND Corporation, Santa Monica, California (comanor@rand.org).

but is otherwise unknown. The best lower bound on $\mathbf{E}\, f_0(X)$ is given by the optimal value of the linear optimization problem

$$\begin{array}{ll} \text{minimize} & \mathbf{E}\, f_0(X) \\ \text{subject to} & \mathbf{E}\, f_i(X) = a_i, \quad i = 1, \ldots, m, \end{array}$$

where we optimize over all probability distributions on $\mathbf{R}^n$. The dual of this problem is

$$\begin{array}{ll} \text{maximize} & z_0 + \sum_{i=1}^m a_i z_i \\ \text{subject to} & z_0 + \sum_{i=1}^m z_i f_i(x) \le f_0(x) \text{ for all } x, \end{array} \tag{1.2}$$

and has a finite number of variables z_i, $i = 0, \ldots, m$, but infinitely many constraints. Isii shows that strong duality holds under appropriate constraint qualifications, so we can find a sharp lower bound on $\mathbf{E}\, f_0(X)$ by solving (1.2). The research on generalized Chebyshev inequalities in the 1960s focused on problems for which (1.2) can be solved analytically.

Isii's formulation is also useful for numerical computation of Chebyshev bounds. In fact the constraints in (1.2) are equivalent to a single constraint

$$g(z_0, \ldots, z_m) \stackrel{\Delta}{=} \sup_x \left(z_0 + \sum_{i=1}^m z_i f_i(x) - f_0(x) \right) \le 0.$$

The function $g : \mathbf{R}^{m+1} \to \mathbf{R}$ is convex, but in general difficult to evaluate, so solving (1.2) is usually a very hard computational problem. In this paper we consider a special case for which (1.2) reduces to a semidefinite programming problem that can be solved efficiently.

The recent development of interior-point methods for nonlinear convex optimization, and semidefinite programming in particular, has revived the interest in generalized Chebyshev inequalities and related moment problems. Bertsimas and Sethurama [BS00], Bertsimas and Popescu [BP05], Popescu [Pop05] and Lasserre [Las02] discuss various types of generalized Chebyshev bounds that can be computed by semidefinite programming. Other researchers, including Nesterov [Nes00], Genin, Hachez, Nesterov and Van Dooren [GHNV03], and Faybusovich [Fay02] have also explored the connections between different classes of moment problems and semidefinite programming.

Outline of the paper. The main result is given in §2, where we present two equivalent semidefinite programs (SDPs) with optimal values equal to the best lower bound on $\mathbf{Prob}(X \in C)$, where C is defined as in (1.1), given the first two moments of the distribution. We also show how to compute a distribution that attains the bound. These SDPs can be derived from Isii's semi-infinite linear programming formulation, combined with a non-trivial linear algebra result known as the S-procedure in control theory [BV04, Appendix B]. Our goal in this paper is to present a simpler and constructive proof based only on (finite-dimensional) convex duality. The theorem is illustrated with a simple example in §3. A geometrical interpretation is given in §4. Some applications and possible extensions are discussed in §5. The appendix summarizes the key definitions and results of semidefinite programming duality theory. More background on semidefinite programming can be found in the books [NN94, WSV00, BTN01, BV04].

Notation. $\mathbf{S}^n$ will denote the set of symmetric $n \times n$ matrices; $\mathbf{S}^n_+$ the set of symmetric positive semidefinite $n \times n$ matrices. For $X \in \mathbf{S}^n$, we write $X \succeq 0$ if X is positive semidefinite, and $X \succ 0$ if X is positive definite. The trace of a matrix X is denoted $\mathbf{tr}\, X$. We use the standard inner product $\mathbf{tr}(XY)$ on $\mathbf{S}^n$.

2. Probability of a set defined by quadratic inequalities. The main result of the paper is as follows. Let C be defined as in (1.1), with $A_i \in \mathbf{S}^n$, $b_i \in \mathbf{R}^n$, and $c_i \in \mathbf{R}$. For $\bar{x} \in \mathbf{R}^n$, $S \in \mathbf{S}^n$ with $S \succeq \bar{x}\bar{x}^T$, we define $\mathbf{P}(C, \bar{x}, S)$ as

$$\mathbf{P}(C, \bar{x}, S) = \inf\{\mathbf{Prob}(X \in C) \mid \mathbf{E}\, X = \bar{x},\ \mathbf{E}\, XX^T = S\},$$

where the infimum is over all probability distributions on $\mathbf{R}^n$.

The optimal values of the following two SDPs are equal to $\mathbf{P}(C, \bar{x}, S)$.

1. (Upper bound SDP)

$$\begin{array}{ll} \text{minimize} & 1 - \sum_{i=1}^m \lambda_i \\ \text{subject to} & \mathbf{tr}(A_i Z_i) + 2b_i^T z_i + c_i \lambda_i \geq 0, \quad i = 1, \ldots, m \\ & \sum_{i=1}^m \left[\begin{array}{cc} Z_i & z_i \\ z_i^T & \lambda_i \end{array}\right] \preceq \left[\begin{array}{cc} S & \bar{x} \\ \bar{x}^T & 1 \end{array}\right] \\ & \left[\begin{array}{cc} Z_i & z_i \\ z_i^T & \lambda_i \end{array}\right] \succeq 0, \quad i = 1, \ldots, m. \end{array} \tag{2.1}$$

   The variables are $Z_i \in \mathbf{S}^n$, $z_i \in \mathbf{R}^n$, and $\lambda_i \in \mathbf{R}$, for $i = 1, \ldots, m$.

2. (Lower bound SDP)

$$\begin{array}{ll} \text{maximize} & 1 - \mathbf{tr}(SP) - 2q^T \bar{x} - r \\ \text{subject to} & \left[\begin{array}{cc} P - \tau_i A_i & q - \tau_i b_i \\ (q - \tau_i b_i)^T & r - 1 - \tau_i c_i \end{array}\right] \succeq 0, \quad i = 1, \ldots, m \\ & \tau_i \geq 0, \quad i = 1, \ldots, m \\ & \left[\begin{array}{cc} P & q \\ q^T & r \end{array}\right] \succeq 0. \end{array} \tag{2.2}$$

   The variables are $P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, $r \in \mathbf{R}$, and $\tau_i \in \mathbf{R}$, for $i = 1, \ldots, m$.

In the remainder of this section we prove this result using semidefinite programming duality. The proof can be summarized as follows.

- In §2.1 and §2.2 we show that the optimal value of the SDP (2.1) is an upper bound on $\mathbf{P}(C, \bar{x}, S)$.
- In §2.3 we show that the optimal value of the SDP (2.2) is a lower bound on $\mathbf{P}(C, \bar{x}, S)$.
- To conclude the proof, we note that the two SDPs are dual problems, and that the lower bound SDP is strictly feasible. It follows from semidefinite programming duality (see the appendix) that their optimal values are therefore equal.

2.1. Distributions that satisfy an averaged quadratic constraint. In this section we prove a linear algebra result that will be used in the constructive proof of the upper bound property in §2.2.

Suppose a random variable $Y \in \mathbf{R}^n$ satisfies

$$\mathbf{E}(Y^T AY + 2b^T Y + c) \geq 0,$$

where $A \in \mathbf{S}^n$, $b \in \mathbf{R}^n$, $c \in \mathbf{R}$. Then there exists a discrete random variable X, with $K \leq 2n$ possible values, that satisfies

$$X^T AX + 2b^T X + c \geq 0, \qquad \mathbf{E}\, X = \mathbf{E}\, Y, \qquad \mathbf{E}\, XX^T \preceq \mathbf{E}\, YY^T.$$

If we denote the moments of Y as $Z = \mathbf{E}\, YY^T$ and $z = \mathbf{E}\, Y$, we can state this result more specifically as follows. Suppose $Z \in \mathbf{S}^n$ and $z \in \mathbf{R}^n$ satisfy

$$Z \succeq zz^T, \qquad \mathbf{tr}(AZ) + 2b^T z + c \geq 0.$$

Then there exist vectors $v_i \in \mathbf{R}^n$ and scalars $\alpha_i \geq 0$, $i = 1, \ldots, K$, with $K \leq 2n$, such that

$$v_i^T A v_i + 2b^T v_i + c \geq 0, \quad i = 1, \ldots, K,$$

and

$$\sum_{i=1}^K \alpha_i = 1, \qquad \sum_{i=1}^K \alpha_i v_i = z, \qquad \sum_{i=1}^K \alpha_i v_i v_i^T \preceq Z. \tag{2.3}$$

Proof. We distinguish two cases, depending on the sign of $\lambda \stackrel{\Delta}{=} z^T Az + 2b^T z + c$. If $\lambda \geq 0$, we can simply choose $K = 1$, $v_1 = z$, and $\alpha_1 = 1$. If $\lambda < 0$, we start by factoring $Z - zz^T$ as

$$Z - zz^T = \sum_{i=1}^n w_i w_i^T,$$

for example, using the eigenvalue decomposition. (We do not assume that the w_i's are independent or nonzero.) We have

$$0 \leq \mathbf{tr}(AZ) + 2b^T z + c = \sum_{i=1}^n w_i^T A w_i + z^T Az + 2b^T z + c = \sum_{i=1}^n w_i^T A w_i + \lambda,$$

and because $\lambda < 0$, at least one of the terms $w_i^T A w_i$ in the sum must be positive. Assume the first r terms are positive, and the last $n - r$ are negative or zero. Define

$$\mu_i = w_i^T A w_i, \quad i = 1, \ldots, r.$$

We have $\mu_i > 0$, $i = 1, \ldots, r$, and

$$\sum_{i=1}^r \mu_i = \sum_{i=1}^r w_i^T A w_i \geq \sum_{i=1}^n w_i^T A w_i \geq -\lambda.$$

For $i = 1, \ldots, r$, let β_i and β_{i+r} be the positive and negative roots of the quadratic equation

$$\mu_i \beta^2 + 2w_i^T (Az + b)\beta + \lambda = 0.$$

The two roots exist because $\lambda < 0$ and $\mu_i > 0$, and they satisfy

$$\beta_i \beta_{i+r} = \lambda / \mu_i.$$

We take $K = 2r$, and, for $i = 1, \ldots, r$,

$$v_i = z + \beta_i w_i, \qquad \alpha_i = \frac{\mu_i}{(1 - \beta_i/\beta_{i+r})(\sum_{k=1}^r \mu_k)},$$

and

$$v_{i+r} = z + \beta_{i+r} w_i, \qquad \alpha_{i+r} = -\alpha_i \beta_i / \beta_{i+r}.$$

By construction, the vectors v_i satisfy $v_i^T A v_i + 2b^T v_i + c = 0$. It is also clear that $\alpha_i > 0$ and $\alpha_{i+r} > 0$ (since $\mu_i > 0$ and $\beta_i/\beta_{i+r} < 0$). Moreover

$$\sum_{i=1}^{2r} \alpha_i = \sum_{i=1}^{r} \alpha_i (1 - \beta_i/\beta_{i+r}) = \sum_{i=1}^{r} \frac{\mu_i}{\sum_{k=1}^{r} \mu_k} = 1.$$

Next, we note that $\alpha_i \beta_i + \alpha_{i+r}\beta_{i+r} = \alpha_i(\beta_i - (\beta_i/\beta_{i+r})\beta_{i+r}) = 0$, and therefore

$$\sum_{i=1}^{K} \alpha_i v_i = \sum_{i=1}^{r} (\alpha_i(z + \beta_i w_i) + \alpha_{i+r}(z + \beta_{i+r} w_i)) = z.$$

Finally, using the fact that $\beta_i \beta_{i+r} = \lambda/\mu_i$ and $\sum_{i=1}^{r} \mu_i \geq -\lambda$, we can prove the third property in (2.3):

$$\begin{aligned}
\sum_{i=1}^{K} \alpha_i v_i v_i^T &= \sum_{i=1}^{r} \left(\alpha_i (z + \beta_i w_i)(z + \beta_i w_i)^T + \alpha_{i+r}(z + \beta_{i+r} w_i)(z + \beta_{i+r} w_i)^T\right) \\
&= \sum_{i=1}^{2r} \alpha_i z z^T + \sum_{i=1}^{r} (\alpha_i \beta_i + \alpha_{i+r}\beta_{i+r})(z w_i^T + w_i z^T) \\
&\quad + \sum_{i=1}^{r} (\alpha_i \beta_i^2 + \alpha_{i+r}\beta_{i+r}^2) w_i w_i^T \\
&= z z^T + \sum_{i=1}^{r} \alpha_i (\beta_i^2 - \beta_i \beta_{i+r}) w_i w_i^T \\
&= z z^T + \sum_{i=1}^{r} \frac{\mu_i}{(1 - \beta_i/\beta_{i+r}) \sum_{k=1}^{r} \mu_k} (\beta_i^2 - \beta_i \beta_{i+r}) w_i w_i^T \\
&= z z^T + \sum_{i=1}^{r} \frac{\mu_i}{\sum_{k=1}^{r} \mu_k} (-\beta_i \beta_{i+r}) w_i w_i^T \\
&= z z^T + \sum_{i=1}^{r} \frac{-\lambda}{\sum_{k=1}^{r} \mu_k} w_i w_i^T \\
&\preceq z z^T + \sum_{i=1}^{r} w_i w_i^T \\
&\preceq Z.
\end{aligned}$$

2.2. Upper bound property. Assume Z_i, z_i, λ_i satisfy the constraints in the SDP (2.1), with $\sum_{i=1}^{m} \lambda_i < 1$. (We will return to the case $\sum_{i=1}^{m} \lambda_i = 1$.) We show that there exists a random variable X with

$$\mathbf{E}\, X = \bar{x}, \qquad \mathbf{E}\, X X^T = S, \qquad \mathbf{Prob}(X \in C) \leq 1 - \sum_{i=1}^{m} \lambda_i.$$

Hence,

$$\mathbf{P}(C, \bar{x}, S) \leq 1 - \sum_{i=1}^{m} \lambda_i.$$

Proof. Without loss of generality we assume that the first k coefficients λ_i are nonzero, and the last $m-k$ coefficients are zero. Using the result of §2.1, and the first and third constraints in (2.1), we can define k independent discrete random variables X_i that satisfy

$$X_i^T A_i X_i + 2b_i^T X_i + c_i \geq 0, \qquad \mathbf{E}\, X_i = z_i/\lambda_i, \qquad \mathbf{E}\, X_i X_i^T \preceq Z_i/\lambda_i \tag{2.4}$$

for $i = 1, \ldots, k$. From the second constraint in (2.1) we see that

$$\sum_{i=1}^k \lambda_i \begin{bmatrix} \mathbf{E}\, X_i X_i^T & \mathbf{E}\, X_i \\ \mathbf{E}\, X_i^T & 1 \end{bmatrix} \preceq \sum_{i=1}^k \begin{bmatrix} Z_i & z_i \\ z_i^T & \lambda_i \end{bmatrix} \preceq \begin{bmatrix} S & \bar{x} \\ \bar{x}^T & 1 \end{bmatrix},$$

so if we define S_0, $\bar{x}_0$ as

$$\begin{bmatrix} S_0 & \bar{x}_0 \\ \bar{x}_0^T & 1 \end{bmatrix} = \frac{1}{1 - \sum_{i=1}^k \lambda_i} \left(\begin{bmatrix} S & \bar{x} \\ \bar{x}^T & 1 \end{bmatrix} - \sum_{i=1}^k \lambda_i \begin{bmatrix} \mathbf{E}\, X_i X_i^T & \mathbf{E}\, X_i \\ \mathbf{E}\, X_i^T & 1 \end{bmatrix} \right),$$

then $S_0 \succeq \bar{x}_0 \bar{x}_0^T$. This means we can construct a discrete random variable X_0 with $\mathbf{E}\, X_0 = \bar{x}_0$ and $\mathbf{E}\, X_0 X_0^T = S_0$, for example, as follows. Let $S_0 - \bar{x}_0 \bar{x}_0^T = \sum_{i=1}^r w_i w_i^T$ be a factorization of $S_0 - \bar{x}_0 \bar{x}_0^T$. If $r = 0$ we choose $X_0 = \bar{x}_0$. If $r > 0$, we define

$$X_0 = \begin{cases} \bar{x}_0 + \sqrt{r} w_i & \text{with probability } 1/(2r) \\ \bar{x}_0 - \sqrt{r} w_i & \text{with probability } 1/(2r). \end{cases}$$

It is easily verified that X_0 satisfies $\mathbf{E}\, X_0 = \bar{x}_0$ and $\mathbf{E}\, X_0 X_0^T = S_0$.

To summarize, we have defined $k+1$ independent random variables $X_0, \ldots, X_k$ that satisfy $X_i^T A_i X_i + 2b_i^T X_i + c_i \geq 0$ for $i = 1, \ldots, k$, and

$$\sum_{i=0}^k \lambda_i \begin{bmatrix} \mathbf{E}\, X_i X_i^T & \mathbf{E}\, X_i \\ \mathbf{E}\, X_i^T & 1 \end{bmatrix} = \begin{bmatrix} S & \bar{x} \\ \bar{x}^T & 1 \end{bmatrix}, \tag{2.5}$$

where $\lambda_0 = 1 - \sum_{i=1}^k \lambda_i$. Now consider the random variable X with the mixture distribution

$$X = X_i \text{ with probability } \lambda_i \text{ for } i = 0, \ldots, k.$$

From (2.5), X satisfies $\mathbf{E}\, X = \bar{x}$, $\mathbf{E}\, XX^T = S$. Furthermore, since $X_1, \ldots, X_k \notin C$, we have $\mathbf{Prob}(X \in C) \leq 1 - \sum_{i=1}^m \lambda_i$, and therefore $1 - \sum_{i=1}^m \lambda_i$ is an upper bound on $\mathbf{P}(C, \bar{x}, S)$.

It remains to consider the case in which Z_i, z_i, and λ_i are feasible in (2.1) with $\sum_{i=1}^m \lambda_i = 1$. Define

$$\tilde{Z}_i = (1-\epsilon) Z_i, \qquad \tilde{z}_i = (1-\epsilon) z_i, \qquad \tilde{\lambda}_i = (1-\epsilon)\lambda_i, \quad i = 1, \ldots, m,$$

where $0 < \epsilon < 1$. These values are also feasible, with $\sum_i \tilde{\lambda}_i = 1 - \epsilon$, so we can apply the construction outlined above and construct a random variable X with $\mathbf{E}\, X = \bar{x}$, $\mathbf{E}\, XX^T = S$, and $\mathbf{Prob}(x \in C) \leq \epsilon$. This is true for any ϵ with $0 < \epsilon < 1$. Therefore $\mathbf{P}(C, \bar{x}, S) = 0$.

2.3. Lower bound property. Suppose P, q, r, and $\tau_1, \ldots, \tau_m$ are feasible in (2.2). Then

$$1 - \mathbf{tr}(SP) - 2q^T \bar{x} - r \leq \mathbf{P}(C, \bar{x}, S).$$

Proof. The constraints of (2.2) imply that, for all x,

$$x^T Px + 2q^T x + r \geq 1 + \tau_i(x^T A_i x + 2b_i^T x + c_i), \quad i = 1, \dots, m,$$

and $x^T Px + 2q^T x + r \geq 0$. Therefore

$$x^T Px + 2q^T x + r \geq 1 - \mathbf{1}_C(x) = \begin{cases} 1 & x \notin C \\ 0 & x \in C, \end{cases}$$

where $\mathbf{1}_C(x)$ denotes the 0-1 indicator function of C. Hence, if $\mathbf{E}\, X = \bar{x}$, and $\mathbf{E}\, XX^T = S$, then

$$\begin{aligned} \mathbf{tr}(SP) + 2q^T \bar{x} + r &= \mathbf{E}(X^T PX + 2q^T X + r) \\ &\geq 1 - \mathbf{E}\, \mathbf{1}_C(X) \\ &= 1 - \mathbf{Prob}(X \in C). \end{aligned}$$

This shows that $1 - \mathbf{tr}(SP) - 2q^T \bar{x} - r$ is a lower bound on $\mathbf{Prob}(X \in C)$.

3. Example. In simple cases, the two SDPs can be solved analytically, and the formulation can be used to prove some well-known inequalities. As an example, we derive an extension of the Chebyshev inequality known as Selberg's inequality [KS66, page 475],

Suppose $C = (-1, 1) = \{x \in \mathbf{R} \mid x^2 < 1\}$. We show that

$$\mathbf{P}(C, \bar{x}, s) = \begin{cases} 0 & 1 \leq s \\ 1 - s & |\bar{x}| \leq s < 1 \\ (1 - |\bar{x}|)^2/(s - 2|\bar{x}| + 1) & s < |\bar{x}|. \end{cases} \tag{3.1}$$

This generalizes the Chebyshev inequality

$$\mathbf{Prob}(|X| \geq 1) \leq \min\{1, \sigma^2\},$$

which is valid for random variables X on $\mathbf{R}$ with $\mathbf{E}\, X = 0$ and $\mathbf{E}\, X^2 = \sigma^2$.

Without loss of generality we assume that $\bar{x} \geq 0$. The upper bound SDP for $\mathbf{P}(C, \bar{x}, s)$ is

$$\begin{array}{ll} \text{minimize} & 1 - \lambda \\ \text{subject to} & Z \geq \lambda \\ & 0 \preceq \begin{bmatrix} Z & z \\ z & \lambda \end{bmatrix} \preceq \begin{bmatrix} s & \bar{x} \\ \bar{x} & 1 \end{bmatrix} \end{array}$$

with variables $\lambda, Z, z \in \mathbf{R}$. If $s \geq 1$, we can take $Z = s$, $z = \bar{x}$, $\lambda = 1$, which has objective value zero. If $\bar{x} \leq s < 1$, we can take $Z = s$, $z = \bar{x}$, $\lambda = s$, which provides a feasible point with objective value $1 - s$. Finally, if $\bar{x} > s$, we can verify that the values

$$Z = z = \lambda = \frac{s - \bar{x}^2}{s - 2\bar{x} + 1} = \frac{s - \bar{x}^2}{s - \bar{x}^2 + (\bar{x} - 1)^2}$$

are feasible. They obviously satisfy $Z \geq \lambda$ and the first matrix inequality. They also satisfy the upper bound, since

$$\begin{bmatrix} s - Z & \bar{x} - z \\ \bar{x} - z & 1 - \lambda \end{bmatrix} = \frac{1}{s - 2\bar{x} + 1} \begin{bmatrix} \bar{x} - s \\ 1 - \bar{x} \end{bmatrix} \begin{bmatrix} \bar{x} - s \\ 1 - \bar{x} \end{bmatrix}^T \succeq 0.$$

The objective function evaluated at this feasible point is

$$1 - \lambda = \frac{(1-\bar{x})^2}{s - 2\bar{x} + 1}.$$

This shows that the righthand side of (3.1) is an upper bound on $\mathbf{P}(C, \bar{x}, s)$.

The lower bound SDP is

$$\begin{array}{ll} \text{maximize} & 1 - sP - 2\bar{x}q - r \\ \text{subject to} & \begin{bmatrix} P & q \\ q & r-1 \end{bmatrix} \succeq \tau \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \\ & \tau \geq 0 \\ & \begin{bmatrix} P & q \\ q & r \end{bmatrix} \succeq 0, \end{array}$$

with variables $P, q, r, \tau \in \mathbf{R}$. The values $P = q = \tau = 0$, $r = 1$ are always feasible, with objective value zero. The values $P = \tau = 1$, $r = q = 0$ are also feasible, with objective value $1 - s$. The values

$$\begin{bmatrix} P & q \\ q & r \end{bmatrix} = \frac{1}{(s - 2\bar{x} + 1)^2} \begin{bmatrix} 1 - \bar{x} \\ s - \bar{x} \end{bmatrix} \begin{bmatrix} 1 - \bar{x} \\ s - \bar{x} \end{bmatrix}^T, \qquad \tau = \frac{1 - \bar{x}}{s - 2\bar{x} + 1}$$

are feasible if $s < \bar{x}$, since in that case $\bar{x}^2 \leq s$ implies $\bar{x} < 1$, and hence

$$\tau = \frac{1 - \bar{x}}{s - \bar{x}^2 + (\bar{x} - 1)^2} > 0$$

and

$$\begin{bmatrix} P - \tau & q \\ q & r + \tau - 1 \end{bmatrix} = \frac{(1 - \bar{x})(\bar{x} - s)}{(s - 2\bar{x} + 1)^2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \succeq 0.$$

The corresponding objective value is

$$1 - sP - 2\bar{x}q - r = \frac{(1 - \bar{x})^2}{s - 2\bar{x} + 1}.$$

This proves that the righthand side of (3.1) is a lower bound on $\mathbf{P}(C, \bar{x}, s)$.

4. Geometrical interpretation. Figure 4.1 shows an example in $\mathbf{R}^2$. The set C is defined by three linear inequalities and two nonconvex quadratic inequalities. The moment constraints are displayed by showing $\bar{x} = \mathbf{E}\, X$ (shown as a small circle), and the set

$$\{x \mid (x - \bar{x})^T (S - \bar{x}\bar{x}^T)^{-1} (x - \bar{x}) = 1\}$$

(shown as the dashed ellipse).

The optimal Chebyshev bound for this problem is $\mathbf{P}(C, \bar{x}, S) = 0.3992$. The six heavy dots are the possible values v_i of the discrete distribution computed from the optimal solution of the upper bound SDP. The numbers next to the dots give $\mathbf{Prob}(X = v_i)$, rounded to four decimal places. Since C is defined as an open set, the five points on the boundary are not in C itself, so $\mathbf{Prob}(X \in C) = 0.3992$ for this distribution. The solid ellipse is the level curve

$$\{x \mid x^T P x + 2q^T x + r = 1\}$$

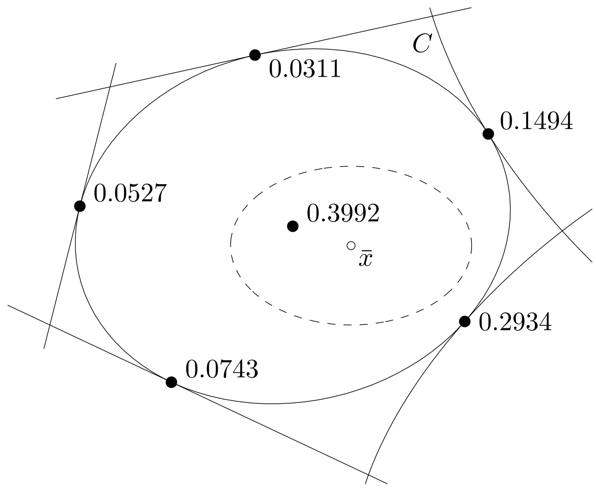

FIG. 4.1. *The set C is the interior of the area bounded by the five solid curves. The dashed ellipse with center $\bar{x}$ is the boundary of the set $\{x \mid (x-\bar{x})^T(S-\bar{x}\bar{x}^T)^{-1}(x-\bar{x}) \le 1\}$. The Chebyshev lower bound on $\mathbf{Prob}(X \in C)$, over all distributions with $\mathbf{E}\,X = \bar{x}$ and $\mathbf{E}\,XX^T = S$, is 0.3992. This bound is sharp, and achieved by the discrete distribution shown with heavy dots. The point with probability 0.3392 lies inside C; the other five points are in the boundary of C, hence not in C. The solid ellipse is the level curve $\{x \mid x^TPx + 2q^Tx + r = 1\}$ where P, q, and r are the optimal solution of the lower bound SDP (2.2).*

where P, q, and r are the optimal solution of the lower bound SDP (2.2).

We notice that the optimal distribution assigns nonzero probability to the points where the ellipse touches the boundary of C, and to its center. This relation between the solutions of the upper and lower bound SDPs holds in general, and can be derived from the optimality conditions of semidefinite programming, as we now show.

Suppose Z_i, z_i, λ_i form an optimal solution of the upper bound SDP, and P, q, r, τ_i are optimal for the lower bound SDP. Consider the set

$$\mathcal{E} = \{x \mid x^TPx + 2q^Tx + r \le 1\},$$

which is an ellipsoid if P is nonsingular. The complementary slackness or optimality conditions for the pair of SDPs (see the appendix) state that

$$\tau_i(\mathbf{tr}(A_iZ_i) + 2b_i^Tz_i + c_i\lambda_i) = 0, \quad i = 1, \ldots, m,$$

$$\mathbf{tr}(PZ_i) + 2q^Tz_i + r\lambda_i = \tau_i(\mathbf{tr}(A_iZ_i) + 2b_i^Tz_i + c_i\lambda_i) + \lambda_i, \quad i = 1, \ldots, m,$$

and

$$\mathbf{tr}(PS) + 2q^T\bar{x} + r = \sum_{i=1}^m(\mathbf{tr}(PZ_i) + 2q^Tz_i + r\lambda_i).$$

Combining the first two conditions gives

$$\mathbf{tr}(PZ_i) + 2q^Tz_i + r\lambda_i = \lambda_i, \quad i = 1, \ldots, m, \tag{4.1}$$

and substituting this in the last condition, we obtain

$$\mathbf{tr}(PS) + 2q^T\bar{x} + r = \sum_{i=1}^m \lambda_i. \tag{4.2}$$

Suppose $\lambda_i > 0$. As we have seen in §2.1, we can associate with Z_i, z_i, λ_i a random variable X_i that satisfies (2.4). Dividing (4.1) by λ_i, we get

$$\mathbf{E}(X_i^T P X_i + 2q^T X_i + r) \le (\mathbf{tr}(PZ_i) + 2q^T z_i + r\lambda_i)/\lambda_i = 1. \tag{4.3}$$

On the other hand,

$$\begin{aligned} X_i^T P X_i + 2q^T X_i + r &\ge X_i^T P X_i + 2q^T X_i + r - \tau_i (X_i^T A_i X_i + 2b_i^T X_i + c_i) \\ &\ge 1, \end{aligned}$$

where the first line follows because $X_i^T A_i X_i + 2b_i^T X_i + c_i \ge 0$ and $\tau_i \ge 0$, and the second line because

$$\left[\begin{array}{cc} P & q \\ q^T & r-1 \end{array}\right] \succeq \tau_i \left[\begin{array}{cc} A_i & b_i \\ b_i^T & c_i \end{array}\right].$$

Combining this with (4.3) we can conclude that

$$X_i^T P X_i + 2q^T X_i + r = 1. \tag{4.4}$$

In other words, if $\lambda_i > 0$ and X_i satisfies (2.4), then X_i lies on the boundary of $\mathcal{E}$.

If $\sum_{i=1}^m \lambda_i < 1$ we can also define a random variable X_0 that satisfies (2.5), and hence

$$\begin{aligned} &(1 - \sum_{i=1}^m \lambda_i)\, \mathbf{E}(X_0^T P X_0 + 2q^T X_0 + r) \\ &= \mathbf{tr}(PS) + 2q^T \bar{x} + r - \sum_{i=1}^m \lambda_i\, \mathbf{E}(X_i^T P X_i + 2q^T X_i + r) \\ &= \sum_{i=1}^m \lambda_i - \sum_{i=1}^m \lambda_i \\ &= 0, \end{aligned}$$

i.e., $\mathbf{E}(X_0^T P X_0 + 2q^T X_0 + r) = 0$. (The second step follows from (4.2) and (4.4).) On the other hand, $X_0^T P X_0 + 2q^T X_0 + r \ge 0$ for all X_0, so we can conclude that

$$X_0^T P X_0 + 2q^T X_0 + r = 0,$$

i.e., X_0 is at the center of $\mathcal{E}$.

5. Conclusion. Generalized Chebyshev inequalities find applications in stochastic processes [PM05], queueing theory and networks [BS00], machine learning [LEBJ02], and communications. The probability of correct detection in a communication or classification system, for example, can often be expressed as the probability that a random variable lies in a set defined by linear or quadratic inequalities. The technique presented in this paper can therefore be used to find lower bounds on the probability of correct detection, or, equivalently, upper bounds on the probability of error. The bounds obtained are the best possible, over all distributions with given first and second order moments, and are efficiently computed using semidefinite programming algorithms. From the optimal solution of the SDPs, the worst-case distribution can be established as described in §2.2.

In practical applications, the worst-case distribution will often be unrealistic, and the corresponding bound overly conservative. Improved bounds can be computed by further restricting the allowable distributions. The lower bound SDP in §2, for example, can be extended to incorporate higher order or polynomial moment constraints [Las02, Par03, BP05], or additional constraints on the distribution such as unimodality [Pop05]. In contrast to the case studied here, however, the resulting Chebyshev bounds are in general not sharp.

REFERENCES

[BP05] D. Bertsimas and I. Popescu. Optimal inequalities in probability theory: a convex optimization approach. *SIAM J. on Optimization*, 15(3):780–804, 2005.

[BS00] D. Bertsimas and J. Sethuraman. Moment problems and semidefinite optimization. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming*, chapter 16, pages 469–510. Kluwer Academic Publishers, 2000.

[BTN01] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization. Analysis, Algorithms, and Engineering Applications*. Society for Pure and Applied Mathematics, 2001.

[BV04] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004. `www.stanford.edu/~boyd/cvxbook`.

[Fay02] L. Faybusovich. On Nesterov's approach to semi-infinite programming. *Acta Applicandae Mathematicae*, 74:195–215, 2002.

[GHNV03] Y. Genin, Y. Hachez, Y. Nesterov, and P. Van Dooren. Optimization problems over positive pseudopolynomial matrices. *SIAM Journal on Matrix Analysis and Applications*, 25(1):57–79, 2003.

[Isi59] K. Isii. On a method for generalizations of Tchebycheff's inequality. *Annals of the Institute of Statistical Mathematics*, 10:65–88, 1959.

[Isi63] K. Isii. On sharpness of Tchebycheff-type inequalities. *Annals of the Institute of Statistical Mathematics*, 14:185–197, 1963.

[Isi64] K. Isii. Inequalities of the types of Chebyshev and Cramér-Rao and mathematical programming. *Annals of The Institute of Statistical Mathematics*, 16:277–293, 1964.

[KS66] S. Karlin and W. Studden. *Tchebycheff Systems: With Applications in Analysis and Statistics*. Wiley-Interscience, 1966.

[Las02] J. Lasserre. Bounds on measures satisfying moment conditions. *The Annals of Applied Probability*, 12(3):1114–1137, 2002.

[LEBJ02] G. Lanckriet, L. El Ghaoui, C. Bhattacharyya, and M. Jordan. A robust minimax approach to classification. *Journal of Machine Learning Research*, 3:555–582, 2002.

[MO60] A. Marshall and I. Olkin. Multivariate Chebyshev inequalities. *Annals of the Mathematical Statistics*, 31:1001–1014, 1960.

[Nes00] Y. Nesterov. Squared functional systems and optimization problems. In J. Frenk, C. Roos, T. Terlaky, and S. Zhang, editors, *High Performance Optimization Techniques*, pages 405–440. Kluwer Academic Publishers, 2000.

[NN94] Y. Nesterov and A. Nemirovsky. *Interior-point polynomial methods in convex programming*, volume 13 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, 1994.

[Par03] P. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming Series B*, 96:293–320, 2003.

[PM05] C. Pandit and S. Meyn. Extremal distributions and worst-case large-deviation bounds. *To appear in Stochastic Processes and their Application*, 2005.

[Pop05] I. Popescu. A semidefinite programming approach to optimal moment bounds for convex classes of distributions. *Mathematics of Operations Research*, 50(3):632–657, 2005.

[WSV00] H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors. *Handbook of Semidefinite Programming*, volume 27 of *International Series in Operations Research and Management Science*. Kluwer Acadmemic Publishers, Boston, MA, 2000.

Appendix A. This appendix summarizes the definition and duality theory of semidefinite programming.

Let $\mathcal{V}$ be a finite-dimensional real vector space, with inner product $\langle u, v\rangle$. Let

$$\mathcal{A} : \mathcal{V} \to \mathbf{S}^{l_1} \times \mathbf{S}^{l_2} \times \cdots \times \mathbf{S}^{l_L}, \qquad \mathcal{B} : \mathcal{V} \to \mathbf{R}^r$$

be linear mappings, where we identify $\mathbf{S}^{l_1} \times \cdots \times \mathbf{S}^{l_L}$ with the space of symmetric block-diagonal matrices with L diagonal blocks of dimensions l_i, $i = 1, \ldots, L$. Suppose $c \in \mathcal{V}$, $D = (D_1, D_2, \ldots, D_L) \in \mathbf{S}^{l_1} \times \cdots \times \mathbf{S}^{l_L}$, and $d \in \mathbf{R}^r$ are given. The optimization problem

$$\begin{array}{ll} \text{minimize} & \langle c, y\rangle \\ \text{subject to} & \mathcal{A}(y) + D \preceq 0 \\ & \mathcal{B}(y) + d = 0 \end{array}$$

with variable $y \in \mathcal{V}$ is called a *semidefinite programming problem* (SDP). The problem is often expressed as

$$\text{(A.1)} \qquad \begin{array}{ll} \text{minimize} & \langle c, y\rangle \\ \text{subject to} & \mathcal{A}(y) + S + D = 0 \\ & \mathcal{B}(y) + d = 0 \\ & S \succeq 0, \end{array}$$

where $S \in \mathbf{S}^{l_1} \times \cdots \times \mathbf{S}^{l_L}$ is an additional slack variable.

The dual SDP associated with (A.1) is defined as

$$\text{(A.2)} \qquad \begin{array}{ll} \text{maximize} & \mathbf{tr}(DZ) + d^T z \\ \text{subject to} & \mathcal{A}^{\text{adj}}(Z) + \mathcal{B}^{\text{adj}}(z) + c = 0 \\ & Z \succeq 0, \end{array}$$

where

$$\mathcal{A}^{\text{adj}} : \mathbf{S}^{l_1} \times \cdots \times \mathbf{S}^{l_L} \to \mathcal{V}, \qquad \mathcal{B}^{\text{adj}} : \mathbf{R}^r \to \mathcal{V}$$

denote the adjoints of $\mathcal{A}$ and $\mathcal{B}$. The variables in the dual problem are $Z \in \mathbf{S}^{l_1} \times \cdots \times \mathbf{S}^{l_L}$, and $z \in \mathbf{R}^r$. We refer to Z as the dual variable (or multiplier) associated with the constraint $\mathcal{A}(y) + D \preceq 0$, and to z as the multiplier associated with the equality constraint $\mathcal{B}(y) + d = 0$.

The *duality gap* associated with primal feasible y, S and a dual feasible Z is defined as

$$\mathbf{tr}(SZ).$$

It is easily verified that the duality gap is equal to the difference between the primal and dual objective functions evaluated at y, S, and Z:

$$\mathbf{tr}(SZ) = \langle c, y\rangle - \mathbf{tr}(DZ) - d^T z.$$

It is also clear that the duality gap is nonnegative, since $S \succeq 0$, $Z \succeq 0$. It follows that the optimal value of the primal problem (A.1) is greater than or equal to the optimal value of the dual problem (A.2). We say *strong duality* holds if the optimal values are in fact equal. It can be shown that a sufficient condition for strong duality is that the primal or the dual problem is strictly feasible.

If strong duality holds, then y, S, Z, z are optimal if and only if they are feasible and the duality gap is zero:

$$\mathbf{tr}(SZ) = 0.$$

The last condition is referred to as *complementary slackness*.